



Radiometric and Geometric Enhancement of UAV-Based Urban Orthophoto Mosaic Using Real-ESRGAN Super-Resolution

MohammadReza Tavakoli¹ , Mohammad SaadatSeresht² , Nasehe Jamshidpour³ , Ali Eftekhari⁴

1. Corresponding author, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: mohammadtavakoli@ut.ac.ir
2. School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: msaadat@ut.ac.ir
3. School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: njamshidpour@ut.ac.ir
4. School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran. Email: ali.eftekhari79@ut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 2025-07-13
Received in revised form 2025-08-15
Accepted 2025-08-19
Published online 2025-10-19

Keywords:
UAV Photogrammetry,
Super-Resolution,
Orthophotomosaics,
Generative Adversarial
Networks (GANs),
Real-ESRGAN

ABSTRACT

This study evaluates the effectiveness of Real-ESRGAN, a deep learning-based super-resolution method, in improving the radiometric and geometric quality of UAV imagery for urban applications. While UAV photogrammetry facilitates the generation of 3D models and orthophotos, its limited spatial resolution restricts accuracy in detailed urban analysis. Super-resolution techniques, particularly those based on deep learning, offer a promising solution by reconstructing finer spatial details from low-resolution inputs.

High-resolution images were reconstructed from UAV-based aerial photographs using the Real-ESRGAN model. These outputs were compared against bicubic interpolation and original datasets to assess visual, radiometric, and geometric improvements. Photogrammetric products, including orthoimage mosaics and 3D mesh models, were generated from each image type. Standard quality metrics (e.g., RMSE, ERGAS, SAM, GRMSE) were used for evaluation.

Real-ESRGAN substantially outperformed bicubic interpolation in radiometric quality, showing improvements of 57.76% in RMSE, 100% in ERGAS, and 56.59% in SAM. It also improved geometric accuracy in derived products, as confirmed by the statistical (z-score) and practical (Cohen's d) analyses, with the largest practical effect observed in the Z-direction and 3D RSS of mesh reconstructions, indicating substantial and practically meaningful error reduction.

The findings demonstrate that Real-ESRGAN can effectively enhance both visual quality and spatial accuracy of UAV-derived imagery and photogrammetric products. However, slight geometric inconsistencies in raw SR images suggest a trade-off between perceptual enhancement and geometric fidelity. Future research should explore geometry-aware super-resolution models that integrate spatial constraints and training strategies suited for geospatial applications.

Cite this article: Tavakoli, M., SaadatSeresht, M., Jamshidpour, N., & Eftekhari, A. (2025). Radiometric and Geometric Enhancement of UAV-Based Urban Orthophoto Mosaic Using Real-ESRGAN Super-Resolution, *Earth Observation and Geomatics Engineering*, Volume 8, Issue 2, Pages 119-132. <http://doi.org/10.22059/eoge.2025.398074.1181>



© The Author(s).

Publisher: University of Tehran.

DOI: <http://doi.org/10.22059/eoge.2025.398074.1181>

1. Introduction

In recent years, the demand for high-precision spatial data in urban areas has grown significantly (Yu & Fang, 2023). This demand is driven by the complexity of urban planning, infrastructure management, and sustainable development. In particular, detailed spatial analysis is required to address challenges such as traffic congestion, environmental pollution, and resource allocation (Rotilio, 2019). As a result, there has been a growing focus on utilizing advanced technologies and methods that produce spatial products with higher geometric and radiometric accuracy (Kovanič et al., 2023). Aerial imagery, particularly when integrated with UAV-based photogrammetry, has become a key source for generating high-resolution 3D models, Orthophotomosaics, and maps that support spatial analyses (Lamsters et al., 2020).

Historically, spatial analysis relied on analog aerial images, which served as the primary source of geospatial data for decades (Muhmad Kamarulzaman et al., 2023). With technological advancements, these analog images were gradually replaced by digital imagery. In recent years, the emergence of unmanned aerial vehicles (UAVs) has transformed urban data acquisition by collecting high-resolution aerial imagery with greater flexibility and cost-effectiveness (Hu & Minner, 2023). This advancement in data collection has improved both the speed of production and the accuracy of spatial models (Lee et al., 2024).

UAV-based photogrammetry has become particularly important in the context of smart city development, where precise spatial information is essential for urban planning, geographic analysis, resource monitoring, and infrastructure management (Mohamed et al., 2020).

However, one of the main challenges in using UAV-derived photogrammetric data is the limitations in the radiometric and geometric quality of the images. Especially in urban environments with complex features, images captured by UAVs can suffer from noise, low resolution, and geometric distortions, which can reduce the accuracy and reliability of subsequent spatial analyses. Therefore, improving image quality has become a key research area in photogrammetry and aerial imaging (Nex et al., 2022). Traditional methods such as interpolation techniques (e.g., nearest neighbor, bilinear, bicubic), noise reduction algorithms, contrast enhancement, and edge reconstruction using filters like Gaussian, median, or Laplacian have been

commonly used. However, these classical approaches face significant limitations. They often fail to reconstruct fine details or preserve complex spatial structures, and in many cases, they lead to blurred, distorted, or artificial results, ultimately reducing the accuracy of photogrammetric applications.

Among the various approaches developed for this purpose, super-resolution techniques have emerged as one of the most promising (Haris et al., 2017; Matsuoka & Fukue, 2020; Panagiotopoulou et al., 2023). These methods aim to reconstruct high-resolution images from low-resolution inputs, thereby providing more detailed and clearer data for urban analysis.

Traditionally, super-resolution methods were based on reconstructing images from a set of images taken with short time or spatial intervals. However, recent advancements in deep learning and neural networks have transformed these techniques, enabling the improvement of image quality from a single input image. Particularly in the last decade, Generative Adversarial Networks (GANs) and models like Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) have garnered significant attention. These models, leveraging advanced deep learning algorithms, allow for the reconstruction of images with higher precision and detail, which is especially beneficial in applications where image clarity is crucial, such as urban planning and 3D modeling.

Implementing these techniques in UAV photogrammetry, particularly for image enhancement, eliminates the need for lower flight altitudes often associated with data redundancy or larger focal lengths, which weaken the network stability. Ultimately, these methods can help reduce costs and improve efficiency in data collection. Despite these advancements, limitations still exist, such as the quality of UAV sensors and non-ideal conditions that can reduce image accuracy. In this context, methods like ESRGAN can significantly mitigate these limitations, leading to higher-quality products in both radiometric and geometric terms.

This paper aims to evaluate the improvement in the radiometric and geometric quality of spatial products derived from UAV photogrammetry using the ESRGAN method. A comparative analysis will be conducted between the original images, enhanced images, and the resulting spatial products such as point clouds, meshes, and orthophoto mosaics. The main objective of this study is to assess the impact of this method on improving the accuracy

and resolution of UAV-based products, as well as reducing errors in spatial data. The findings of this research could serve as an effective solution to overcome the challenges of quality in UAV photogrammetry data in urban areas, particularly in fields such as urban planning, precise mapping, and 3D modeling

Image super-resolution (SR) techniques can be categorized broadly as Single-Image Super-Resolution (SISR) or Multi-Image Super-Resolution (MISR) methods. SISR reconstructs high-resolution (HR) images from single low-resolution (LR) inputs and is widely used in medical imaging, surveillance, and remote sensing. It is popular because of its simplicity and ability to be applied to a single frame. Still, single image super-resolution (SISR) is an ill-posed problem because high-frequency details are lost, making their recovery challenging. This enables AI-based SISR techniques to outperform classical methods (Bee Lim et al., 2017).

MISR, on the other hand, uses multiple LR images of the same scene viewed from different angles or at different points in time. It fuses these images to generate HR outputs through complementary information, which is applicable in video super-resolution, astronomy, and photogrammetry. However, MISR methods require precise image alignment and are computationally heavy, which restricts their use in dynamic or noisy environments (Christian Ledig et al., 2017).

SISR is the main focus of this review, and as such, our proposed ESRGAN method is included in this category. We explore classical and AI-based approaches and review the respective novelties, advantages, and limitations.

A key gap identified in the current body of research is that most super-resolution studies have focused solely on improving the visual quality of images, without adequately evaluating their geometric accuracy and applicability in geospatial decision-making. Investigations reveal that while these algorithms may visually enhance images, they can introduce significant distortions in the geometric structure, which compromises the reliability of the resulting spatial products. This becomes critical in high-precision applications, where minor inaccuracies may lead to considerable errors in final outputs and spatial analyses. The present study aims to address this limitation by not only evaluating visual improvements but also systematically assessing the geometric integrity of super-resolved UAV images. The research introduces a novel, application-oriented approach that integrates deep learning techniques with photogrammetric principles to propose a new pathway for improving both the visual and spatial quality of aerial imagery.

The main scientific contributions of this research can be summarized as follows, aiming to address current gaps in the literature and advance the practical application of deep learning in UAV-based photogrammetry:

- The study evaluates the performance of Real-ESRGAN in enhancing both the visual and quantitative quality of UAV images compared to classical approaches.

- It assesses the effect of image super-resolution on the positional accuracy of end products using reference images and a set of low-resolution (LR), bicubic interpolated (BI) high-resolution, Real-ESRGAN super-resolution (SR) images.
- The methodology incorporates real and complex urban models along with precise analytical tools to measure geometric accuracy and spatial fidelity.
- The research integrates deep learning with rigorous spatial evaluation to offer a cost-effective, time-efficient, and accurate approach for enhancing photogrammetric products.

Paper structure explanation

2. Single Image super-resolution methods

Single Image Super-Resolution (SISR) techniques aim to reconstruct a high-resolution image from a single low-resolution input, without relying on additional frames or viewpoints. These methods have gained significant attention due to their practicality in real-world scenarios where only one image is available, such as UAV photogrammetry. Over the years, SISR approaches have evolved from traditional interpolation and reconstruction-based algorithms to advanced deep learning architectures that can recover fine textures and structural details. This section reviews both conventional and AI-based SISR methods, highlighting their principles, strengths, and limitations in the context of spatial data enhancement.

2.1. Conventional SISR methods

Conventional super-resolution in the days before deep learning consisted of interpolation-based, reconstruction-based, and example-based methods. These techniques were straightforward and intuitive to implement, yet they were limited in recovering fine details and handling high-frequency textured content.

2.1.1. Interpolation-based methods

These methods, such as bilinear, bicubic, and nearest-neighbor interpolation, use neighboring pixel values to calculate missing high-frequency details by taking an average. These methods are computationally light and easy to apply but fail to recreate fine textures, leading to blurriness and artifacts. Some key studies in this domain include a systematic review of interpolation techniques from the University of Malaysia Perlis (Wang et al., 2018), and a study by HTX's S&COE team comparing interpolation versus deep learning-based methods (Baghel et al., 2023). Introductory materials are also available from general resources such as "Image Scaling" by Wikipedia contributors (Mei et al., 2020).

2.1.2. Reconstruction-based methods

These methods treat SR as an optimization problem and impose priors such as smoothness, sparsity, or edge continuity to constrain the solution space. These methods

are better than interpolation since they also consider assumptions about image properties, but they perform worse in complex real-world settings. Notable works include (Kamasak et al., 2005)'s approach of interpolating and then estimating sub-pixel shift, and (Protter et al., 2009), who used self-similarity-based image reconstruction techniques. (Zhang et al., 2012) showcased progress and issues with the application of such methods to heterogeneous datasets.

2.1.3. Example-based methods

These approaches can employ dictionaries of LR-HR pairs to synthesize the missing high-frequency content. They utilize prior data to improve resolution and act as an intermediary between interpolation and learning-based methods. For instance, (Zhang et al., 2011) applied dictionary learning to medical imaging for CT image reconstruction. (Wang et al., 2015) offered an extensive review of dictionary-based and learning-based SR methods. Additionally, (Bedi & Agarwal, 2015) investigated example-based methods and highlighted that well-curated dictionaries are essential for achieving effective SR.

2.2. DL-based SISR methods

The advent of AI-based methods has transformed SISR, as algorithms can now learn complex LR-to-HR mappings using vast amounts of data. Such techniques may include CNN-based, GAN-based, or Transformer-based methods.

2.2.1. CNN-based methods

Many SISR models employ CNNs, which are the foundation of modern image processing pipelines. One prominent architecture is the Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) (B. Lim et al., 2017). EDSR optimized the ResNet structure by removing batch normalization layers, which improved performance and reduced computational cost. This yielded improved spatial information preservation and state-of-the-art performance on benchmark datasets. However, EDSR faces challenges in reconstructing fine details in highly complex images.

2.2.2. GAN-based methods

Recent advances in SISR include the use of GANs, which have made remarkable progress by focusing on perceptual quality. SRGAN (C. Ledig et al., 2017) comprises two principal components: a generator network and a discriminator network.

2.2.3. Transformer-based methods

Recently, Transformer-based architecture has become popular in SISR. (Baghel et al., 2023) proposed SRTransGAN, which combines the long-range dependency modelling of transformers with the perceptual realism of GANs. This method excels at capturing relationships across the entire image but is computationally expensive and requires large datasets. (Liang et al., 2021) proposed a

novel approach, Image Super-Resolution Using Cross-Scale Non-Local Attention, to model both local and global dependencies using cross-scale attention mechanisms to achieve high-quality texture restoration. However, Transformer-based methods are still computationally expensive and highly reliant on large-scale datasets despite their superior performance. requirements and difficulties in working with heterogeneous datasets.

While progress in SISR techniques is evident, they have not yet been widely applied to UAV-based photogrammetry. UAV-derived spatial products, such as orthoimage mosaics (OIM) and 3D models, introduce challenges like varying resolutions and noise. Adaptive AI-based SISR techniques for UAV photogrammetry should be explored in future research, and their usefulness in improving spatial accuracy and recovering minute details should be assessed. Experiments on original images and derived products are essential to validate their practical utility in remote sensing and geospatial applications.

3. Methodology

This section outlines the methodology employed to analyze image data and associated spatial products for super-resolution improvement. As it is shown in Figure 1, the process begins with data collection, where raw UAV images are obtained and their associated spatial products are generated to serve as the ground truth for our evaluation. Subsequently, low-resolution images are generated using bilinear resampling techniques. Then, high-resolution images are produced through cubic interpolation. Next, super-resolution images are created using the REAL-ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) model. Following that, spatial products of all three groups of images, including Orthoimage Mosaic (OIM) and 3D meshes, are generated to enrich the dataset. Afterward, a comprehensive comparison is conducted between raw images, OIM, and 3D meshes against the ground truth. Finally, the quality assessment is performed to evaluate both the radiometric and geometric accuracy of the products. This evaluation employs predefined metrics such as GRMSE, RRMSE, ERGAS, SAM, UIQI, CC, and PSNR.

3.1. Generating Low-Resolution Images Using Bilinear Resampling

To simulate low-resolution images (Low-Resolution or LR), the high-resolution (High-Resolution or HR) images were down sampled to a lower resolution using 2D bilinear interpolation. This method estimates the value of each new

pixel by averaging the weighted values of neighboring pixels. More precisely, the value of each new pixel $I'(x', y')$ is calculated using the following equation:

$$I'(x', y') = \sum_{i,j} I(x_i, y_j) \cdot w(x', x_i) \cdot w(y', y_j)$$

(1)

Here, $I(x_i, y_j)$ represents the surrounding pixel values, and w are the weights based on the distance between the

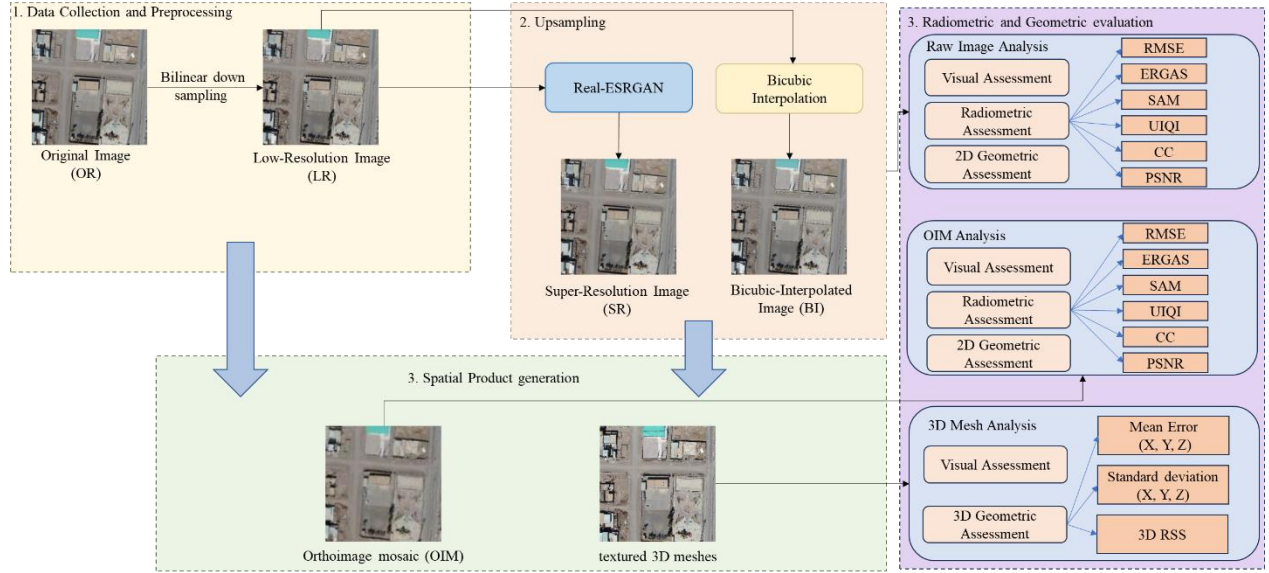


Figure 1. The proposed framework to comprehensive radiometric and geometric evaluation of associated spatial products for super-resolution improvement.

target pixel (x', y') and the original pixels. In this study, the images were down sampled such that their resolution was reduced by a factor of four, and these low-resolution images were then used as input for the REAL-ESRGAN neural network.

3.2. Generating High-Resolution Images Using Bicubic Interpolation

To evaluate the performance of the widely used conventional interpolation technique, cubic interpolation, in improving image quality and the quality of associated spatial products, we applied this method to all images in the dataset. By comparing the results of cubic interpolation with those obtained using the deep learning Real-ESRGAN method, we aimed to quantify the effectiveness of each approach in enhancing image quality and the corresponding spatial products.

3.3. Image Super-Resolution Using Real-ESRGAN

The evolution of generative adversarial networks (GANs) for image super-resolution began with the introduction of SRGAN (Super-Resolution GAN), which was the first to combine perceptual loss and adversarial learning to produce high-resolution images from low-resolution inputs. SRGAN utilized a ResNet-based generator with batch normalization layers and a standard discriminator network, enabling it to generate photo-realistic textures. However, it

often introduced unnatural artifacts and failed to preserve fine details in smooth regions or edges.

To address these limitations, ESRGAN (Enhanced SRGAN) (Wang et al., 2018), introduced several key architectural improvements. It replaced standard residual blocks with Residual-in-Residual Dense Blocks (RRDBs), removing batch normalization to improve training stability and image fidelity. ESRGAN also adopted a relativistic discriminator, which evaluates how much more realistic a generated image is compared to a real one, rather than simply classifying it as real or fake. These changes led to better texture reconstruction, sharper edges, and fewer artifacts than SRGAN, though ESRGAN still relied on synthetic LR-HR image pairs and did not generalize well to real-world degradation patterns.

To overcome this, Real-ESRGAN was developed as a robust extension of ESRGAN for handling real-world degradations such as noise, blur, and compression artifacts (Wang et al., 2021). As shown in Figure 2, it introduces a U-Net-based discriminator for better perceptual quality assessment and uses a high-order degradation model to simulate more realistic low-quality training data. Furthermore, it can be trained with both synthetic and unpaired real-world images. This makes Real-ESRGAN particularly suitable for photogrammetric products like orthophotos and 3D textured models, where image sharpness, geometric consistency, and radiometric clarity

are critical. Its ability to enhance low-quality UAV imagery while preserving structural detail directly improves the spatial accuracy and interpretability of derived geospatial datasets.

A distinctive feature of REAL-ESRGAN is its incorporation of a perceptual loss function, which leverages features extracted from pre-trained deep networks. This loss function helps the network not only optimize traditional quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR) but also improve the perceptual quality of the output images as judged by human observers.

Algorithm 1: REAL-ESRGAN

Input:

Low-resolution image (LR)

Initialize:

- Generator Network G
- Discriminator Network D

1. Define Generator Network G:

- Apply initial convolution to extract basic features from LR image

For each Residual Dense Block in G:

- Apply several convolutional layers
- Use dense connections
- Use residual connections

End

- Apply Super-Resolution Module to upscale image

2. Define Discriminator Network D:

- Apply convolutional layers to extract texture/detail features from image
- Apply fully connected layers to compute realism score

4. Define Perceptual Loss:

- Extract high-level features from pre-trained network
- Compute feature similarity between generated HR and ground truth

5. Adversarial Training Loop:
While until convergence:

- Update Generator G:
 - Generate HR image from LR input
 - Compute perceptual loss and adversarial loss
 - Backpropagate and update G's weights
- Update Discriminator D:
 - Receive real HR image and generated HR image
 - Compute loss to distinguish between real and fake
 - Backpropagate and update D's weights

End

Return High-resolution image (HR) from the Generator

The Real-ESRGAN model was adapted using a two-stage process: an initial pre-training with the RealESRNet architecture for pixel-level reconstruction, followed by fine-tuning with a GAN-based framework incorporating perceptual loss from VGG19, L1 loss (weight 0.01), and structural fidelity terms. The training dataset combined

synthetically degraded images and real UAV imagery from urban environments, ensuring both generalized and domain-specific performance optimization.

3.4. Spatial Products Generation

Following the generation of high-resolution images using the super-resolution model, these outputs were utilized to derive key spatial products essential for photogrammetric and geospatial workflows. One primary application was the construction of 3D mesh models, where the enhanced spatial details contributed to the creation of more geometrically accurate and visually realistic surface representations. These models provide critical structural context and are particularly beneficial in urban modeling, topographic reconstruction, and change detection analyses.

Additionally, orthoimage mosaics (OIMs) were produced from the super-resolved images. These orthorectified mosaics offer high radiometric and geometric fidelity, making them suitable for spatial analysis, thematic mapping, and land monitoring applications. The use of enhanced-resolution inputs in generating OIMs ensured improved sharpness in feature boundaries, better alignment across overlapping scenes, and greater overall interpretability. These spatial products also served as a basis for quantitatively evaluating the benefits of the super-resolution process in terms of both radiometric enhancement and geometric precision.

4. Experiment

This section presents the analysis of the results obtained from evaluating the quality of the reconstructed images. For a comprehensive assessment, it is necessary not only to compare the reconstructed images themselves but also to examine the quality of the derived photogrammetric products, including 3D mesh models and orthophoto mosaics. The section begins with a description of the data collection process.

4.1. Study Area and Imaging Sensor

The study area is in the city of Rafsanjan, situated in Kerman Province in southeastern Iran, which is shown in Figure 3. Rafsanjan lies within a semi-arid climate zone, characterized by clear skies and favorable lighting conditions, making it an ideal environment for aerial imaging and photogrammetric analysis. The region's geographic and climatic features support high-precision data acquisition, offering optimal conditions for capturing high-quality aerial imagery.

The aerial images used in this study were acquired using a DJI Phantom 4 Pro V2 drone equipped with a 20-megapixel camera. The images were captured at an altitude of 90 meters above ground level during the winter of 2021. The camera's sensor produced images with a resolution of 5472×3648 pixels, providing detailed and high-quality data suitable for subsequent super-resolution processing

and spatial product generation.

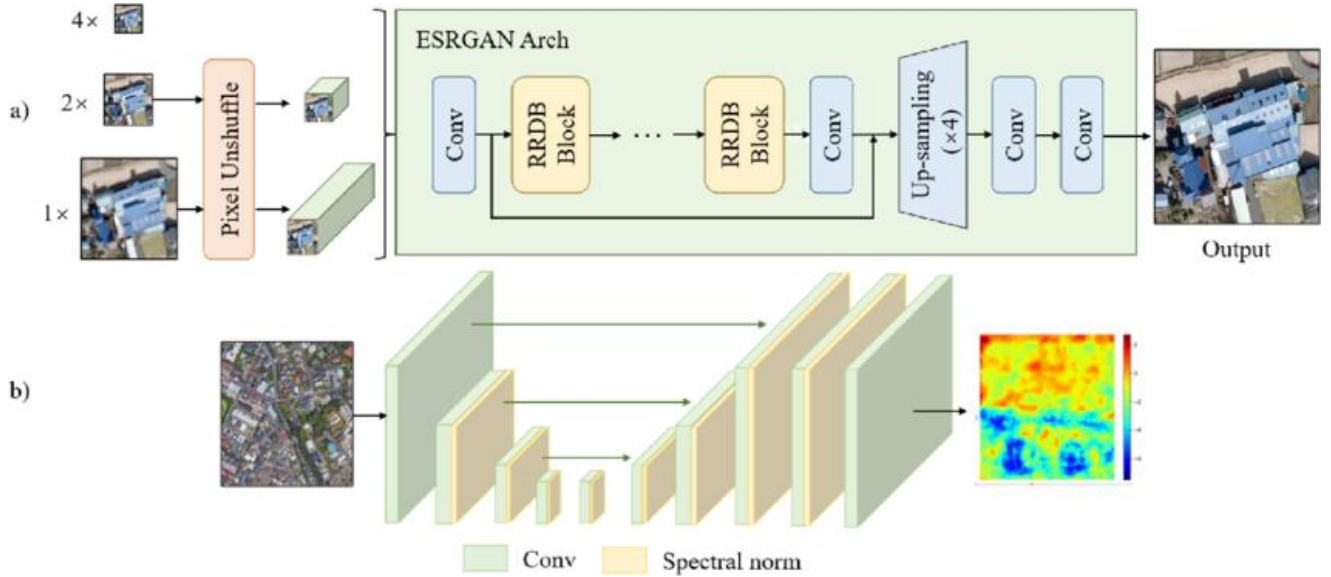


Figure 2. Architecture of the Real-ESRGAN network. (a) Generator; (b) U-net discriminator with spectral normalization (Wang et al., 2021)

4.2. Data Acquisition and Preparation

In this study, aerial images were acquired using drone-based imaging under optimal weather and lighting conditions to ensure high image quality. These high-resolution images formed the primary dataset for the subsequent super-resolution and photogrammetric analyses.



Figure 3. The study area (Rafsanjan, Iran).

reducing their resolution to one-quarter of the original size. These down sampled images served as the low-resolution (LR) inputs for two distinct super-resolution approaches: (1) a bicubic interpolation (BI) model and (2) the Real-ESRGAN deep neural network model. Both methods produced enhanced images with a spatial resolution approximately four times greater than the LR inputs. As a result, three datasets were prepared: the original low-resolution set, the BI-enhanced images, and the REAL-ESRGAN-enhanced images. These datasets were subsequently used to generate photogrammetric outputs, including 3D point clouds, 3D mesh models, and orthoimage mosaics (OIMs), which served as the basis for evaluating spatial and radiometric improvements.

Several challenges arose during the data preparation phase. These included the selection of suitable training image pairs for the deep learning model, as well as the tuning and optimization of the REAL-ESRGAN network to ensure consistent output quality. Moreover, photogrammetric considerations such as image coverage, acquisition geometry, and atmospheric clarity played a crucial role in ensuring accurate data collection. A particularly important challenge involved the reassignment of geospatial information to the enhanced images, as the network processing pipeline did not preserve geotag metadata. This issue was resolved by integrating the

As a preprocessing step, the original high-resolution images were down sampled using bilinear interpolation,

original geospatial metadata, thereby ensuring correct alignment and georeferencing of the final spatial products.

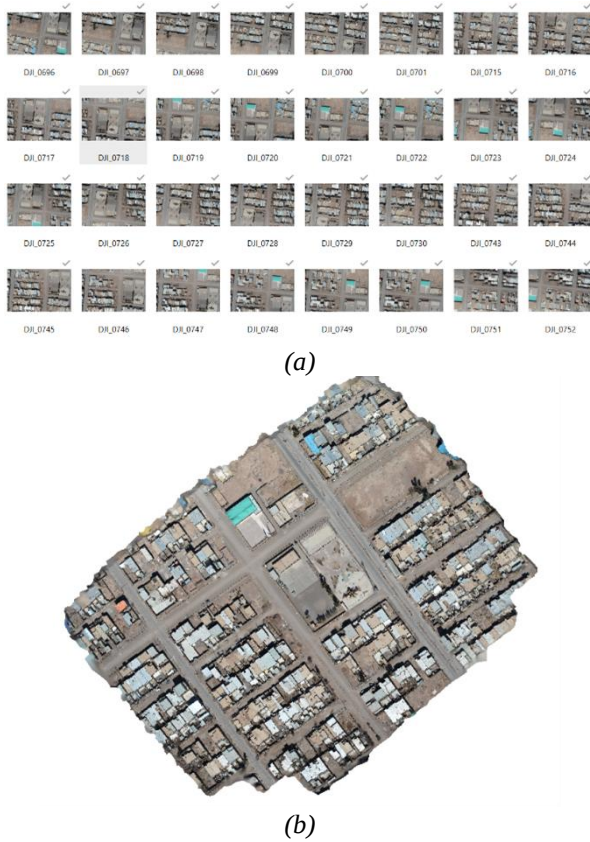


Figure 4. (a) 45 UAV image dataset for evaluation, (b) reconstructed city area as ground truth

4.3. Quality Assessment

To evaluate the accuracy and reliability of the generated outputs, a comprehensive quality assessment was performed by comparing the raw images, orthorectified imagery maps (OIMs), and 3D mesh models against ground truth references. In all evaluation stages, the original high-resolution UAV images (OR), acquired under optimal weather and lighting conditions, served as the ground truth. Low-resolution (LR) images were generated from the OR dataset by applying a $4\times$ downsampling using bilinear interpolation. These LR images were used as inputs for both bicubic interpolation (BI) and Real-ESRGAN (SR) processing. The evaluation framework addressed both radiometric and geometric characteristics of the datasets, which are critical in determining the quality and usability of photogrammetric products. Three categories of analysis, visual, radiometric, and geometric were applied across all derived products to ensure robust and holistic assessment.

Radiometric evaluation was conducted using widely adopted image quality metrics, including Root Mean Square Error (RMSE), Relative RMSE (RRMSE), Error relative global dimensionless synthesis (ERGAS), Spectral Angle Mapper (SAM), Universal Image Quality Index (UIQI), Correlation Coefficient (CC), and Peak Signal-to-Noise

Ratio (PSNR). These metrics were applied to representative raw images and OIMs, comparing them to the original high-resolution dataset as the reference. Geometric evaluation was performed using the Ground Root Mean Square Error (GRMSE). This metric was applied in 2D for raw images and OIMs and in 3D for the reconstructed 3D mesh models, allowing precise assessment of positional accuracy.

Visual inspection was conducted through qualitative comparison of selected urban features, particularly buildings, in both raw images and OIMs. This method provided insights into perceptual improvements, especially regarding edge sharpness and structural fidelity, resulting from the super-resolution process. The evaluation was based on three image sources: low-resolution images, bicubic-interpolated images, and images enhanced via Real-ESRGAN super-resolution.

To quantify the comparative performance of the methods, the percentage improvement (IMP) or degradation in each quality metric was calculated between SR and BI methods, relative to the LR baseline. The improvement metric is computed as follows:

$$IMP = \left(\frac{SR - BI}{BI - LR} \right) \times 100 \quad (2)$$

For the visual and geometric assessments, representative urban features were selected, including various types of residential and commercial buildings, major road networks, complex intersections, and open spaces. These features were chosen to capture the structural diversity and varying spatial complexities of the study area. For the radiometric assessment, 38 images were selected to cover diverse urban and semi-urban scenes, ensuring statistical representativeness.

For geometric assessments of raw images and orthophoto mosaics, five representative image scenes were chosen to reflect variations in building density, structural complexity, and open spaces, with 30 well-distributed corresponding points measured for each scene. In the 3D mesh analysis, Cloud-to-Mesh distances were computed from all vertices of the reconstructed mesh, with measurements averaged separately for the X, Y, and Z coordinates to ensure statistically robust and spatially comprehensive accuracy measurements.

4.4. Raw Image Analysis

4.4.1. Visual assessment

The visual assessment was conducted by five independent reviewers to minimize bias. The original high-resolution (OR) images, 4-times down sampled low-resolution (LR) images, and super-resolved images produced by the bilinear interpolation (BI), and Real-ESRGAN model (SR) were presented side-by-side for direct comparison. Reviewers evaluated each image using a standardized scoring form covering eight quality indicators: sharpness, noise level, color fidelity, contrast, presence of artifacts, structural similarity, naturalness, and fine detail preservation. Each indicator was rated on a five-point scale, ensuring

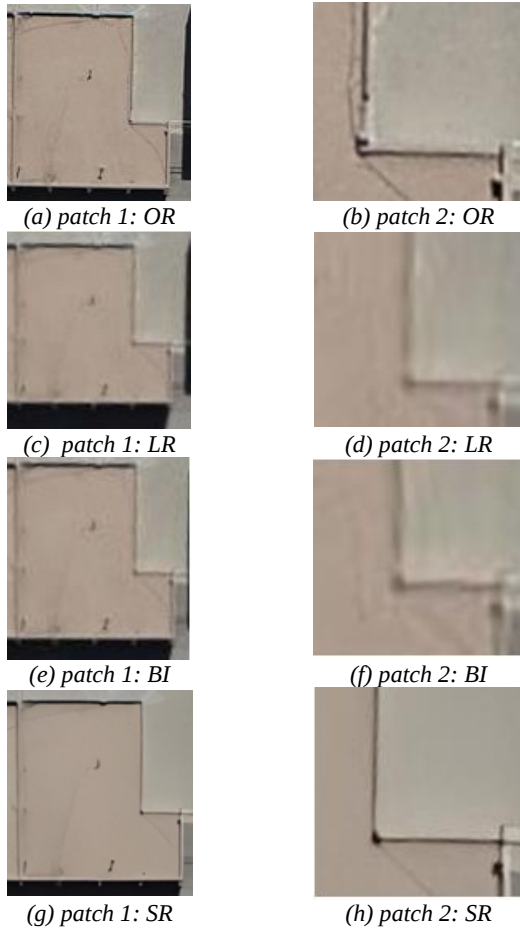


Figure 5. Visual comparison of two sample patches from UAV imagery across different processing methods. (a–b) Original high-resolution images (OR); (c–d) low-resolution inputs (LR); (e–f) bicubic interpolation results (BI); (g–h) super-resolved outputs using Real-ESRGAN (SR).

consistency and comparability across image types. Table 1, presenting the visual assessment scores for one of the evaluated images.

As illustrated in Figure 5, the LR image is low quality which details are missing details. The BI image has better quality compared to LR, but there is degradation in the edges. The SR image is very close to OR and has the best visual quality.

Table 1. Raw Images Visual Analysis

Criterion	LR vs OR	BI vs OR	SR vs OR
Sharpness	1	3	4
Noise	1	3	5
Color Fidelity	2	3	4
Contrast	3	3	5
Artifacts	1	3	5
SSIM	1	2	5
Naturalness	2	3	4
Detail	1	2	3
Avg.	1.5714	2.8571	4.2857

4.4.2. Radiometric assessment

To further analyze the quality of the reconstructed images, a radiometric evaluation was conducted on 38 images. In this assessment, the RMSE, ERGAS, SAM, UIQI, CC, and PSNR metrics were computed to quantify the extent of quality improvement in the reconstructed images compared to the original ones. The results of this analysis are presented in Table 2.

The computed IMP values in Table 2 provide a quantitative measure of the relative improvement achieved by the Real-ESRGAN (SR) method over the bicubic interpolation (BI), with respect to the low-resolution (LR) baseline. Notably, the SR method consistently outperforms BI across all radiometric criteria. For RMSE, ERGAS, and SAM—which are error-based metrics where lower values indicate better performance, SR achieves improvements of 57.76%, 100%, and 56.59%, respectively, confirming substantial reduction in radiometric distortion. For quality-enhancing metrics such as UIQI, CC, and PSNR, SR again demonstrates significant gains with IMP values of 90.44%, 62.63%, and 61.12%, respectively. These results clearly show that Real-ESRGAN is very effective at improving the accuracy and visual quality of UAV images, performing much better than traditional interpolation methods.

Table 2. Raw Image Radiometric Analysis

Radiometric Criteria	LR	BI	SR	IMP
RMSE	0.05914	0.05637	0.05477	57.76
ERGAS	0.00023	0.00022	0.00021	100.0
SAM	0.10368	0.09875	0.09596	56.59
UIQI	0.95310	0.95791	0.96226	90.43
CC	0.95608	0.95996	0.96239	62.62
PSNR	72.7070	73.1213	73.3745	61.11

4.4.2. 2D Geometric assessment

In addition to radiometric evaluation, a geometric accuracy assessment was conducted to analyze the spatial consistency of the reconstructed images (Table 3). For this purpose, five pairs of different images were selected, and the GRMSE (Geometric Root Mean Square Error) was calculated for each pair. This metric provides a quantitative measure of geometric deviations by comparing corresponding spatial points across whole corresponding images.

Although Real-ESRGAN enhances the visual quality of UAV images, GRMSE results show that it does not always improve geometric accuracy. In some areas, low-resolution (LR) images actually exhibit lower geometric error than SR outputs. This reveals that perceptual improvements do not necessarily correspond to better spatial alignment—an important limitation in applications requiring high geometric fidelity. The observed differences are typically sub-pixel and may seem negligible; however, in high-precision photogrammetric applications, even such small discrepancies can impact results.

The underlying cause is that deep models like Real-ESRGAN reconstruct images based on learned features, rather than preserving exact pixel geometry. This can introduce subtle misalignments, especially in complex or repetitive structures, increasing GRMSE despite better visual clarity. Therefore, while super-resolution enhances visual and radiometric properties, it may come at the cost of geometric precision, a trade-off that must be considered in photogrammetric applications.

Table 3. Raw Image 2D Geometric Analysis

Image	LR	BI	SR
Image 1	0.261	0.269	0.962
Image 2	0.344	0.458	1.231
Image 3	0.266	0.394	1.630
Image 4	0.270	0.588	1.733
Image 5	0.246	0.335	1.454
Avg.	0.2774	0.4088	1.402

4.5. Orthoimage Mosaics (OIM) Analysis

4.5.1. OIM visual assessment

The visual quality of orthoimage mosaics generated from each image category (OR, LR, BI, SR) was evaluated through side-by-side comparisons, which are shown in Figure 6. The assessment focused on visual artifacts in urban areas, where LR images showed significant blurring and loss of fine detail, especially in shadowed regions and complex structures. In contrast, the orthoimage mosaics generated from SR images displayed clearer textures and sharper reconstructed details. Visual comparisons revealed that HRE outputs closely resembled the reference orthophotos (OR), demonstrating the Real-ESRGAN model's effectiveness in enhancing visual fidelity and preserving structural integrity in photogrammetric products.

4.5.2. OIM radiometric assessment

Following the visual evaluations on OIMS, the analysis extended to radiometric assessment on the orthophotomosaics generated from LR, BI, and SR images. These outputs were assessed in comparison with the reference orthophoto (OR) based on quantitative radiometric metrics to evaluate the reconstruction quality and alignment with the ground truth. The given results in Table 4 indicated that orthophotomosaics generated from SR images consistently outperformed those from LR and BI sources. The SR outputs were closer to the reference in most metrics, reflecting superior radiometric fidelity.

The numerical results in Table 4 show that the Real-ESRGAN method outperforms bicubic interpolation across most radiometric criteria. The most significant improvements are observed in SAM and RMSE, with IMP values of 123.47% and 115.584%, respectively, indicating a notable reduction in spectral and reconstruction errors.

UIQI also shows a strong improvement of 66.388%, suggesting

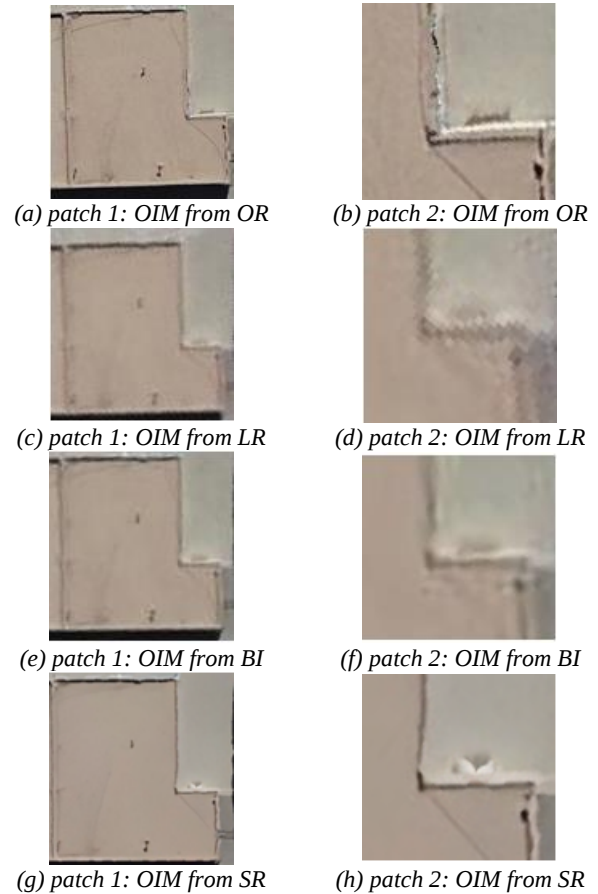


Figure 6. Visual comparison of two sample patches from orthoimage mosaics (OIMs) across different processing methods. (a–b) Original high-resolution images (OR); (c–d) low-resolution inputs (LR); (e–f) bicubic interpolation results (BI); (g–h) super-resolved outputs using Real-ESRGAN (SR).

better structural image quality in the SR results. PSNR and CC exhibit moderate gains of 28.40% and 19.284%, respectively, reflecting enhanced signal clarity and correlation with the original images. The only metric with no improvement is ERGAS, which remains unchanged at 0%, indicating equal performance between SR and BI in this case. Overall, the numerical trends confirm that Real-ESRGAN offers substantial radiometric benefits over traditional interpolation methods in the generation of orthoimage mosaics.

Table 4. Orthoimage Mosaic (OIM) Radiometric Analysis

Radiometric Criteria	LR	BI	SR	IMP
RMSE	0.1205	0.1128	0.1039	115.584
ERGAS	0.00044	0.00040	0.00040	0.0000
SAM	0.2075	0.196	0.1818	123.47
UIQI	0.6029	0.6987	0.7623	66.388
CC	0.6198	0.7121	0.7299	19.284
PSNR	66.596	67.698	68.011	28.40

4.5.3. OIM 2D geometric assessment

To evaluate the geometric accuracy of the orthoimage mosaics generated from images with different resolutions, a 2D positional error analysis was conducted. Using 30 corresponding points per image, the pixel-wise Euclidean distance between the orthoimage from the reference image (OR) and three other input types, low-resolution (LR), bicubic-interpolated (BI), and Real-ESRGAN super-resolved (SR) was calculated. The average errors for five representative image scenes are reported in Table 5.

Table 5. OIM 2D Geometric Analysis

Image	LR	BI	SR
Image 1	2.833	3.426	2.641
Image 2	2.674	3.261	2.394
Image 3	3.109	3.448	2.988
Image 4	2.728	3.117	2.543
Image 5	3.291	3.376	3.043
Avg.	2.9270	3.3256	2.7218

The analysis confirms that orthophoto mosaics generated from LR images suffer from reduced geometric accuracy due to their lower clarity and radiometric quality. This limitation results in noticeable positional discrepancies during the reconstruction of features, particularly in scenes with complex details. As a consequence, the ability of photogrammetric algorithms to extract precise locations and produce geometrically accurate orthophotos is diminished. In contrast, the mosaics derived from HRE images—reconstructed using the Real-ESRGAN model—demonstrate superior geometric fidelity. The numerical results support this conclusion: the average positional error for SR images was 2.7218 pixels, lower than that of LR (2.9270 pixels) and significantly better than BI (3.3256 pixels). These findings indicate that Real-ESRGAN not only preserves texture and structural details more effectively but also enables more accurate extraction of corresponding points and alignment with the reference image.

4.6. 3D Mesh Analysis

4.6.1. 3D mesh visual assessment

The textured 3D meshes generated from each of the image sets (OR, LR, BI, and SR) were visually analyzed and placed side by side using identical views and similar angles to allow for a precise comparative assessment presented in Figure 7.

This analysis focused on the reconstruction quality of urban features with fine details, such as building edges, roof lines, facades, window frames, and height differences between structural elements.

The results showed that meshes derived from LR images had relatively flat surfaces, rough textures, and low quality, with visible fragmentation and poor reconstruction of structurally diverse areas. These meshes failed to accurately preserve elevation information. In contrast, the meshes generated from SR images exhibited higher clarity and continuity, especially along edges, and showed improved

surface smoothness and coherence compared to LR. They were also better at reconstructing complex urban structures. The reference mesh (OR), as expected, demonstrated the highest accuracy in representing urban features and served as the main comparison standard. However, the meshes produced from SR images performed very similarly to the reference and proved to be effective in visualizing urban details.

4.6.2. 3D mesh geometric assessment

In contrast to the 2D geometric evaluations, which used a limited number of corresponding points, the 3D mesh geometric assessment was carried out using all vertices of the reconstructed mesh. The distances between each vertex of the reconstructed mesh and the reference mesh were calculated in CloudCompare® and averaged separately for the X, Y, and Z coordinates, ensuring statistically robust and spatially comprehensive accuracy measurements.

The conducted analysis aimed to evaluate the influence of input image quality on the geometric accuracy of textured 3D mesh models produced through photogrammetric processes. Four distinct image types were assessed: original reference (OR), low-resolution (LR), and super-resolved high-resolution images enhanced via bicubic interpolation (BI) and the Real-ESRGAN method (SR). The geometric accuracy was quantitatively analyzed using Cloud-to-Mesh distances, measured separately in the X, Y, and Z directions, between each reconstructed mesh and the reference mesh.

The results summarized in Table 6 clearly highlight the performance differences across the tested image resolutions. The lowest mean errors relative to the reference were consistently recorded for the SR meshes, indicating enhanced geometric fidelity achieved by the Real-ESRGAN method. Specifically, the SR meshes had mean positional deviations of approximately 0.0023 m in the X direction, 0.0060 m in Y, and 0.1199 m in Z, compared to the significantly higher errors observed in the bicubic-interpolated BI meshes (X: 0.0038 m, Y: 0.0130 m, Z: 0.1975 m) and LR meshes (X: 0.0037 m, Y: 0.0084 m, Z: 0.3153 m). Moreover, standard deviation values, indicative of consistency and stability in mesh reconstruction, further demonstrate superior performance for SR. For instance, the standard deviations for SR were considerably lower (e.g., 0.0830 m in X, 0.1199 m in Y, 0.2403 m in Z) compared to the other approaches, confirming greater uniformity in reconstruction quality.

The last row of Table 6 presents the overall three-dimensional geometric error magnitude computed using the Root Sum of Squares (RSS). This metric combines the mean errors and their standard deviations from the X, Y, and Z directions into a single, representative measure, providing a clear comparison among the three methods (LR, BI, and SR).

Table 6. 3D Mesh Geometric Error (Avg $\pm$ std)

Direction	LR	BI	SR
X	0.0037 $\pm$ 0.1204	0.0038 $\pm$ 0.0900	0.0023 $\pm$ 0.1203
Y	0.0084 $\pm$ 0.0993	0.0130 $\pm$ 0.1270	0.0060 $\pm$ 0.1299
Z	0.1199 $\pm$ 0.3153	0.1975 $\pm$ 0.2403	0.0051 $\pm$ 0.1998
3D RSS	0.1203 $\pm$ 0.3518	0.1980 $\pm$ 0.2863	0.0082 $\pm$ 0.2670

According to the computed RSS values, the SR method significantly outperforms both LR and BI. Specifically, the RSS value for SR (0.0082 $\pm$ 0.2670) is considerably lower compared to LR (0.1203 $\pm$ 0.3518) and BI (0.1980 $\pm$ 0.2863). This indicates that the SR method achieved the smallest overall positional deviation and highest geometric accuracy among the three methods. Additionally, the lower standard deviation of the SR method reflects greater consistency and reliability in reconstructing detailed 3D structures. Conversely, the BI method exhibited the highest mean RSS value, implying substantial geometric distortion introduced by bicubic interpolation compared to the original LR images. These results underscore the superior effectiveness of the Real-ESRGAN super-resolution method (SR) in enhancing geometric fidelity in photogrammetrically generated 3D meshes.

To further investigate the differences in geometric errors between the evaluated methods, we quantified both statistical significance and practical significance using two complementary measures: the z-score test and Cohen's d effect size. The z-score measures how many standard errors separate the means of two groups and is calculated as

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (3)$$

Where $\bar{x}_1$ and $\bar{x}_2$ are the sample means, σ_1 and σ_2 are the standard deviations, and n_1 and n_2 are the sample sizes for the two groups. Under the null hypothesis of equal means, $|z| > 1.96$ indicates statistical significance at the 95% confidence level (two-tailed). Given the very large number of samples in this study, comparable to the number of mesh vertices in the 3D models, it is expected that z-scores can reach extremely high values even for very small absolute differences in the means. This means that while the statistical significance is high for nearly all comparisons, the practical relevance of these differences must be interpreted with caution.

To complement the z-score analysis, Cohen's d was calculated to assess practical significance, which reflects the magnitude of the observed differences in a scale-independent manner. Cohen's d is given by

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p} \quad (4)$$

with the pooled standard deviation s_p defined as

$$s_p = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} \quad (5)$$

Cohen's suggested interpretation thresholds are: small ($0.2 \leq d < 0.5$), moderate ($0.5 \leq d < 0.8$), and large ($d \geq 0.8$). Unlike the z-score, Cohen's d is not directly influenced by sample size, making it a more robust measure of the actual magnitude of the improvement. Cohen's suggested thresholds are: small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d < 0.8$), and large ($d \geq 0.8$).

Table 7. Statistical significance (z-score) and practical significance (Cohen's d) for 3D mesh geometric error comparisons

Direction	BI vs. LR	SR vs. LR	SR vs. BI
X	0.26 (0.0)	3.19 (0.01)	3.87 (0.01)
Y	11.05 (0.04)	5.68 (0.021)	14.92 (0.05)
Z	75.81 (0.28)	119.11 (0.435)	238.44 (0.87)
3D RSS	66.35 (0.24)	98.31 (0.359)	187.77 (0.69)

Table 7 presents the z-scores (outside parentheses) and Cohen's d values (inside parentheses) for each pairwise comparison. The z-scores are extremely large in most cases, indicating statistical significance at the 95% level. However, given the very large number of vertices in the 3D mesh (often exceeding 150,000 points), even minuscule differences in means can yield large z-scores. This inflates statistical significance and can be misleading if interpreted alone.

While z-scores are very large across most comparisons, confirming that improvements are statistically significant, the majority of Cohen's d values are in the small range, indicating that the effect magnitude is modest in practical terms. Notable exceptions are found in the vertical (Z) direction and the aggregated 3D RSS measure for the SR vs. BI comparison, where effect sizes reach the large category ($d = 0.87$ and $d = 0.69$, respectively). These results are consistent with the known sensitivity of vertical accuracy in photogrammetric 3D reconstructions to image resolution and geometric quality. Vertical (Z) estimates depend heavily on parallax measurements, which are more susceptible to improvements in image sharpness and texture detail. Because vertical measurements typically have higher baseline-related errors and noise compared to horizontal (X and Y) measurements, resolution enhancement through super-resolution methods can lead to proportionally greater accuracy gains in Z. This explains why the Z-direction in our results exhibits the largest practical significance, even when horizontal improvements remain statistically significant but practically smaller in magnitude.

Overall, the analysis conclusively demonstrates that employing deep-learning-based super-resolution methods such as Real-ESRGAN not only enhances the visual and textural quality of input images but also significantly improves the geometric accuracy and reliability of resulting 3D mesh models. This outcome emphasizes the considerable potential of advanced image-enhancement methods in

refining photogrammetric products beyond conventional interpolation techniques.

5. Discussion and Conclusion

The experimental results confirm that the Real-ESRGAN model significantly enhances both the visual and radiometric quality of UAV-derived spatial products when compared to conventional bicubic interpolation. Across raw images, orthoimage mosaics (OIMs), and textured 3D meshes, Real-ESRGAN consistently delivers improvements in texture clarity, feature sharpness, and radiometric fidelity, as evidenced by substantial gains in metrics such as RMSE, SAM, UIQI, and PSNR.

In the case of radiometric analysis, Real-ESRGAN achieved improvements of up to 100% in some metrics compared to traditional methods, showing its effectiveness in reducing spectral and reconstruction errors. These enhancements contribute directly to more visually coherent and information-rich images, which can support better interpretation and decision-making in urban remote sensing tasks.

From a geometric perspective, the findings reveal a more subtle outcome. While Real-ESRGAN improved the geometric fidelity of orthoimage mosaics and 3D mesh reconstructions in aggregate, its performance on raw 2D imagery exhibited certain limitations. In some cases, the SR images showed higher GRMSE values than their LR counterparts, indicating that perceptual enhancement may come at the cost of spatial accuracy. This misalignment arises from the nature of deep learning models, which prioritize feature reconstruction over pixel-level geometric consistency. As such, even sub-pixel discrepancies, while visually negligible, can undermine the accuracy of photogrammetric outputs in applications requiring precise spatial measurements.

The results for 3D mesh reconstruction are particularly promising. The SR-derived meshes demonstrated substantial reductions in error magnitude across all dimensions (X, Y, Z) and outperformed both LR and BI methods in terms of consistency and spatial fidelity. While z-score analysis indicated that most differences between methods were statistically significant, partly due to the very large sample size, Cohen's *d* analysis showed that practical significance was concentrated in the vertical (Z) component and the aggregated 3D RSS, where improvements were of moderate to large magnitude, indicating meaningful gains in real-world 3D reconstruction performance. It is critical for urban feature modeling and topographic analyses.

Despite these advances, the trade-off between radiometric enhancement and geometric accuracy remains a central challenge. These sub-pixel geometric discrepancies are primarily caused by the network's feature reconstruction process, which can introduce slight spatial shifts during convolution and upsampling, especially in areas with repetitive or fine structural patterns. While negligible for

visual interpretation, such deviations can accumulate in high-precision photogrammetric workflows, potentially reducing the geometric accuracy of the final products.

Super-resolution networks, including Real-ESRGAN, are not inherently optimized for spatial coherence, which can lead to artifacts in geometric alignment, especially in areas with repetitive or fine structural patterns. This underscores the need for future research on geometry-aware super-resolution architectures that integrate spatial constraints and photogrammetric principles directly into the learning process. For future research, exploring hybrid super-resolution approaches could further refine image quality and spatial accuracy.

Another limitation of this study is that the experiments were conducted on a single dataset from Rafsanjan, Iran, acquired under optimal weather and lighting conditions. While this ensured high-quality input data, it may limit the generalizability of the findings to other environments, such as regions with different climatic conditions, varying urban densities, or UAV platforms with different sensor specifications. Future research should include case studies across diverse environmental settings and sensor configurations to assess the method's performance in broader operational contexts. Furthermore, validation using data from various UAV sensors under different environmental conditions, such as variable lighting, seasonal changes, and adverse weather—would provide a more comprehensive evaluation of Real-ESRGAN's robustness across a wider range of scenarios.

Additionally, integrating high-resolution data, advanced imaging hardware, and deep learning-based enhancement techniques could offer a more robust framework for improving the precision and quality of geospatial products. Such advancements could bridge the gap between visual quality and geometric fidelity, enhancing the reliability of super-resolved imagery in high-precision geospatial applications.

References

- Baghel, N., Dubey, S. R., & Singh, S. K. (2023). SRTransGAN: Image Super-Resolution using Transformer based Generative Adversarial Network. *arXiv preprint arXiv:2312.01999*.
<https://doi.org/10.48550/arXiv.2312.01999>
- Haris, M., Watanabe, T., Fan, L., Widyanto, M. R., & Nobuhara, H. (2017). Superresolution for UAV images via adaptive multiple sparse representation and its application to 3-D reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 4047–4058.
<https://doi.org/https://doi.org/10.1109/tgrs.2017.2687419>
- Hu, D., & Minner, J. (2023). UAVs and 3D City Modeling to Aid Urban Planning and Historic Preservation: A Systematic Review. *Remote Sensing*, 15(23), 5507.
<https://doi.org/10.3390/rs15235507>
- Kamasak, M. E., Bouman, C. A., Morris, E. D., & Sauer, K. (2005). Direct reconstruction of kinetic parameter images from dynamic PET data. *IEEE transactions on medical imaging*, 24(5), 636–650. <https://doi.org/10.1109/TMI.2005.845317>

- Kovanič, L., Topitzer, B., Peťovský, P., Blišťan, P., Gergel'ová, M. B., & Blišťanová, M. (2023). Review of photogrammetric and lidar applications of UAV. *Applied Sciences*, 13(11), 6732. <https://doi.org/10.3390/app13116732>
- Lamsters, K., Karušs, J., Krievāns, M., & Ješkins, J. (2020). High-resolution orthophoto map and digital surface models of the largest Argentine Islands (the Antarctic) from unmanned aerial vehicle photogrammetry. *Journal of Maps*, 16(2), 335–347. <https://doi.org/10.1080/17445647.2020.1748130>
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., & Wang, Z. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, <https://doi.org/10.1109/CVPR.2017.19>
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, <https://doi.org/10.1109/CVPR.2017.19>
- Lee, E., Park, S., Jang, H., Choi, W., & Sohn, H.-G. (2024). Enhancement of low-cost UAV-based photogrammetric point cloud using MMS point cloud and oblique images for 3D urban reconstruction. *Measurement*, 226, 114158. <https://doi.org/10.1016/j.measurement.2024.114158>
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image super-resolution using swin transformer *IEEE International Conference on Computer Vision (ICCV)*, <https://doi.org/10.48550/arXiv.2108.10257>
- Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, <https://doi.org/10.1109/CVPRW.2017.151>
- Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, <https://doi.org/10.1109/CVPRW.2017.151>
- Matsuoka, R., & Fukue, K. (2020). Preliminary Investigation on Possibility of Super Resolution of Uav Orthoimages. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 301–308. <https://doi.org/10.5194/isprs-annals-V-1-2020-301-2020>
- Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T. S., & Shi, H. (2020). Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, <https://doi.org/10.1109/CVPR42600.2020.01156>
- Mohamed, N., Al-Jaroodi, J., Jawhar, I., Idries, A., & Mohammed, F. (2020). Unmanned aerial vehicles applications in future smart cities. *Technological forecasting and social change*, 153, 119293. <https://doi.org/10.1016/j.techfore.2018.05.004>
- Muhmad Kamarulzaman, A. M., Wan Mohd Jaafar, W. S., Mohd Said, M. N., Saad, S. N. M., & Mohan, M. (2023). UAV implementations in urban planning and related sectors of rapidly developing nations: A review and future perspectives for Malaysia. *Remote Sensing*, 15(11), 2845. <https://doi.org/10.3390/rs15112845>
- Nex, F., Armenakis, C., Cramer, M., Cucci, D. A., Gerke, M., Honkavaara, E., Kukko, A., Persello, C., & Skaland, J. (2022). UAV in the advent of the twenties: Where we stand and what is next. *ISPRS journal of photogrammetry and remote sensing*, 184, 215–242. <https://doi.org/10.1016/j.isprsjprs.2022.04.004>
- Panagiotopoulou, A., Grammatikopoulos, L., El Saer, A., Petsa, E., Charou, E., Ragia, L., & Karras, G. (2023). Super-resolution techniques in photogrammetric 3D reconstruction from close-range UAV imagery. *Heritage*, 6(3), 2701–2715. <https://doi.org/10.3390/heritage6030143>
- Protter, M., Elad, M., Takeda, H., & Milanfar, P. (2009). Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, 18(1), 36–51. <https://doi.org/https://doi.org/10.1109/TIP.2008.2008067>
- Rotilio, M. (2019). UAV photogrammetry for resilience management in reconstruction plan of urban historical centres after seismic events. A case study. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprs-archives-XLII-2-W11-55-2019>
- Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. *Proceedings of the IEEE/CVF international conference on computer vision*, <https://doi.org/10.1109/ICCVW54120.2021.00217>
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. *Proceedings of the European conference on computer vision (ECCV) workshops*, https://doi.org/10.1007/978-3-030-11021-5_5
- Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior *IEEE International Conference on Computer Vision (ICCV)*, <https://doi.org/10.1109/ICCV.2015.50>
- Yu, D., & Fang, C. (2023). Urban remote sensing with spatial big data: A review and renewed perspective of urban studies in recent decades. *Remote Sensing*, 15(5), 1307. <https://doi.org/10.3390/rs15051307>
- Zhang, K., Gao, X., Tao, D., & Li, X. (2012). Single image super-resolution with non-local means and steering kernel regression. *IEEE Transactions on Image Processing*, 21(11), 4544–4556. <https://doi.org/https://doi.org/10.1109/TIP.2012.2208977>
- Zhang, Y., Dong, Z., Wu, L., & Wang, S. (2011). A hybrid method for MRI brain image classification. *Expert Systems with Applications*, 38(8), 10049–10053. <https://doi.org/https://doi.org/10.1016/j.eswa.2011.02.012>